Brain response

Authors

Michael Flower[1], Vilija Lomeikaite[2], Peter Holmans[3], Lesley Jones[3], Sarah J. Tabrizi[1] and Darren G. Monckton[2]

1. Dept of Neurodegenerative Disease and Dementia Research Institute, UCL, UK
2. Institute of Molecular, Cell and Systems Biology, University of Glasgow, UK
3. MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, UK

We recently demonstrated that genetic variation in and around exon 1 of *MSH3* is associated with somatic expansion of the CAG•CTG repeat and disease severity in Huntington's disease (HD) and myotonic dystrophy type 1 (DM1) (Flower *et al.*, 2019). Specifically, we revealed that in our well characterised European cohorts (HD, *n* = 218; DM1, *n* = 247), the *MSH3* 9 bp repeat allele, 3a, was associated with reduced *MSH3* expression in blood and brain, with less somatic expansion, and with later age at onset and slower progression. Three non-coding single nucleotide polymorphisms (SNPs) in *MSH3* were in near complete linkage disequilibrium (LD) with the 3a allele, and therefore similarly associated with each phenotype. In addition, after controlling for *MSH3* repeat length, two other SNPs were independently associated with onset. These same *MSH3* exon 1 haplotype associations with both age at onset and somatic instability have recently been independently replicated in HD patients of European ancestry by the latest Genetic Modifiers of HD genome wide association study (*n* = 9,064) (Lee *et al.*, 2019) and by Ciosi *et al.* (2019a) (*n* = 734). These data thus firmly establish that *MSH3* variants are associated with both somatic expansion and disease severity in HD. An obvious question arising, is whether *MSH3* variants are also associated with disease severity in other simple sequence repeat expansion disorders? In Bettencourt *et al.* (2016), we showed that DNA repair variants that influence HD onset are also associated with onset in other CAG repeat expansion polyglutamine diseases, though this was primarily driven by spinocerebellar ataxia type 2 (SCA2) and SCA6.

In their letter, Yau et al. investigated whether the *MSH3* 3a allele was associated with disease severity in 132 spinocerebellar ataxia type 3 (SCA3) and 136 Friedreich's ataxia (FRDA) patients. They found no significant associations. Of course, such a finding may indicate that somatic expansion is not a critical feature of these disorders and/or that *MSH3* does not drive somatic expansion of the CAG repeat in SCA3 or the GAA repeat in FRDA. However, it is interesting to note, that their pattern of age at onset variation with 3a genotype (Figure 1) is not dissimilar to our data (Figure 2, (Flower *et al.*, 2019)), in that two copies of 3a appear to delay onset. Indeed, as the authors acknowledge, it remains highly plausible that the failure of Yau et al. to detect a statistically significant association represents a false negative type II error. In our analysis, the number of 3a alleles accounted for around 5% of variability in onset. With such a small sample, Yau et al.

would have had only around 70% power to detect a similar effect at a significance level of 0.05. Our HD progression score integrated longitudinal motor, cognitive and imaging data and has proved to be a highly sensitive measure of HD severity (Hensman Moss *et al.*, 2017; Ciosi *et al.*, 2019b). In contrast, Yau et al. investigated the effect on disease progression based on a single measure, the SARA score, which was only available for a very small sub-sample of FRDA patients (n = 57). SARA is a relatively insensitive semi-quantitative measure of ataxia (Burk *et al.*, 2013), requiring a sample size over 200 to detect a 50% change in annual progression (Jacobi *et al.*, 2015). Additionally, whilst SCA3 onset correlates with repeat length, SARA score does not (Huang *et al.*, 2017), potentially because of its insensitivity, or because mechanisms underlying onset and progression are distinct.

In addition to the small sample size and insensitivity of the clinical progression measures, technical constraints in the assay and possible population-specific effects may also have limited the ability of Yau et al. to detect an effect for *MSH3* variants. Whereas we unambiguously genotyped all *MSH3* repeat alleles using amplicon sequencing, Yau et al. identified 3a alleles by a combination of fragment length analysis, and by deconvoluting superimposed Sanger sequencing traces in heterozygotes. Notably, Yau et al., revealed a 3a allele frequency in the combined SCA3 and FRDA cohort of 0.35. This was significantly higher than the 0.25 and 0.27 in our European HD and DM1 subjects (Chi-squared = 15.48, $p = 0.0008$). This is likely due to the heterogeneous ancestry of their cohort, which included African, Caribbean, Asian and South American subjects, regional differences in *MSH3* allele frequencies (Nakajima *et al.*, 1995), and/or the inability to resolve 3a alleles from other similar alleles, especially in populations likely to contain additional diversity from that detected in Europe. In addition, whilst our data implicates 3a (Flower *et al.*, 2019), we cannot exclude an effect from other *MSH3* variants in LD with 3a. It is very possible that *MSH3* SNP and repeat LD patterns differ in non-European populations, potentially further confounding the ability of Yau et al., to define an association with 3a in their heterogenous cohort.

In our analyses, we used Illumina sequencing in HD (Ciosi et al.), and, small pool PCR and restriction digest analysis in DM1 (Cumming *et al.*, 2019), to define both the sequence and inherited progenitor allele length of disease-associated CAG•CTG alleles. In contrast, Yau et al., measured the length of the SCA3 and FRDA disease-associated repeats by fragment analysis and took no account of the potential confounding effects of age-dependent somatic expansions. The gross confounding effects of overestimating allele length due to age-dependent somatic expansion on establishing genotype to phenotype correlations in DM1 is well established (Morales *et al.*, 2012). Although the levels of somatic expansion of small HD disease-causing alleles (40 to 50 repeats) in blood DNA are relatively low, permitting the unambiguous identification of the inherited progenitor allele, larger HD alleles (>50 CAG repeats) can show much higher levels of somatic expansion, obviating the unambiguous identification of the progenitor allele (Ciosi *et al.*, 2019b). Given the

much longer average length of SCA3 (mean CAG = 70 repeats) and FRDA (mean GAA ~ 1000) alleles, it seems very likely that determination of the progenitor allele length has been compromised by somatic expansion. Such an effect would introduce a bias in which the length of more somatically unstable alleles are systematically even more overestimated, leading to an underestimation of the degree of variation due to allele length, and thus potentially directly masking the disease accelerating effects of *MSH3* variants that promote somatic expansion. Failure to take account of the effects of somatic expansion in driving measured allele length may also explain the lower proportion of variability in onset observed in SCA3 and FRDA (46% and 36% respectively), relative to the ~60-70% typically observed in HD (Andresen *et al.*, 2007) and DM1 (Morales *et al.*, 2012). Moreover, fragment length analysis sheds no light on the presence or absence of variant repeats that have been shown to have major disease modifying effects in HD (Ciosi *et al.*, 2019b; Lee *et al.*, 2019), DM1 (Braida *et al.*, 2010; Cumming *et al.*, 2019) and FRDA (Cossee *et al.*, 1997; McDaniel *et al.*, 2001; Sakamoto *et al.*, 2001; Pollard *et al.*, 2004), and are known to exist in some SCA3 alleles (Limprasert *et al.*, 1996).

Although it is clear that the expanded SCA3 repeat is somatically unstable (Watanabe *et al.*, 1996; Hashida *et al.*, 1997; Cancel *et al.*, 1998), the degree of somatic instability is lower than in HD and SCA1 (Maciel *et al.*, 1995; Cancel *et al.*, 1998). This is despite the fact that expanded SCA3 alleles are typically much larger (70 to 80 CAG repeats) than in HD and SCA1 (40 to 50 CAG repeats). Indeed, it has been suggested that inherited SCA3 alleles need to be so large to cause disease because they are somatically more stable and need to start closer to the pathogenic threshold in neurons to illicit disease during the lifetime of an individual (Nestor and Monckton, 2011). Likewise, the inherited disease-causing GAA alleles in FRDA are already very large, and although they are somatically unstable (De Biase *et al.*, 2007; Long *et al.*, 2017), the extensive somatic expansions observed in HD (Kennedy *et al.*, 2003) and DM1 (Anvret *et al.*, 1993; Thornton *et al.*, 1994) affected tissues have not been reported. These observations do not preclude a role for further somatic expansion in driving SCA3 and FRDA, but suggest that these two disorders may be less susceptible to somatic expansion than HD and DM1, and effects of *MSH3* variants may therefore be more subtle.

In conclusion, Yau et al. did not challenge our associations between *MSH3* variants and symptomatic variability and somatic expansion in HD and DM1, but rather their extension to SCA3 and FRDA. Their study was small and underpowered to detect changes in onset and progression, and does not preclude the involvement of *MSH3* in other polyglutamine diseases. Indeed, we think it premature to assert that “repeat variant in *MSH3* is not a genetic modifier for spinocerebellar ataxia 3 and Friedreich ataxia”, but agree with the authors that a larger study of repeat disorders will be important to identify common genetic factors.

References

Andresen JM, Gayan J, Djousse L, Roberts S, Brocklebank D, Cherny SS*, et al.* The relationship between CAG repeat length and age of onset differs for Huntington's disease patients with juvenile onset or adult onset. Annals of Human Genetics 2007; 71(Pt 3): 295–301.

Anvret M, Ahlberg G, Grandell U, Hedberg B, Johnson K, Edstrom L. Larger expansions of the CTG repeat in muscle compared to lymphocytes from patients with myotonic dystrophy. Hum Mol Genet 1993; 2: 1397–400.

Bettencourt C, Hensman-Moss D, Flower M, Wiethoff S, Brice A, Goizet C*, et al.* DNA repair pathways underlie a common genetic mechanism modulating onset in polyglutamine diseases. Ann Neurol 2016; 79(6): 983-90.

Braida C, Stefanatos RK, Adam B, Mahajan N, Smeets HJ, Niel F*, et al.* Variant CCG and GGC repeats within the CTG expansion dramatically modify mutational dynamics and likely contribute toward unusual symptoms in some myotonic dystrophy type 1 patients. Hum Mol Genet 2010; 19(8): 1399–412.

Burk K, Schulz SR, Schulz JB. Monitoring progression in Friedreich ataxia (FRDA): the use of clinical scales. J Neurochem 2013; 126 Suppl 1: 118-24.

Cancel G, Gourfinkel-An I, Stevanin G, Didierjean O, Abbas N, Hirsch E*, et al.* Somatic mosaicism of the CAG repeat expansion in spinocerebellar ataxia type 3/Machado-Joseph disease. Hum Mutat 1998; 11(1): 23-7.

Ciosi M, Maxwell A, Cumming SA, Hensman Moss DJ, Alshammari AM, Flower MD*, et al.* A genetic association study of glutamine-encoding DNA sequence structures, somatic CAG expansion, and DNA repair gene variants, with Huntington disease clinical outcomes. EBioMedicine 2019a; 48: 568-80.

Ciosi M, Maxwell A, Cumming SA, Hensman Moss DJ, Alshammari AM, Flower MD*, et al.* A genetic association study of glutamine-encoding DNA sequence structures, somatic CAG expansion, and DNA repair gene variants, with Huntington disease clinical outcomes. EBioMedicine 2019b.

Cossee M, Schmitt M, Campuzano V, Reutenauer L, Moutou C, Mandel JL*, et al.* Evolution of the Friedreich's ataxia trinucleotide repeat expansion: founder effect and premutations. Proceedings of the National Academy of Sciences of the United States of America 1997; 94(14): 7452-7.

Cumming SA, Jimenez-Moreno C, Okkersen K, Wenninger S, Daidj F, Hogarth F*, et al.* Genetic determinants of disease severity in the myotonic dystrophy type 1 OPTIMISTIC cohort. Neurology 2019.

De Biase I, Rasmussen A, Monticelli A, Al-Mahdawi S, Pook M, Cocozza S*, et al.* Somatic instability of the expanded GAA triplet-repeat sequence in Friedreich ataxia progresses throughout life. Genomics 2007; 90(1): 1-5.

Flower M, Lomeikaite V, Ciosi M, Cumming S, Morales F, Lo K*, et al.* MSH3 modifies somatic instability and disease severity in Huntington's and myotonic dystrophy type 1. Brain 2019.

Hashida H, Goto J, Kurisaki H, Mizusawa H, Kanazawa I. Brain regional differences in the expansion of a CAG repeat in the spinocerebellar ataxias: dentatorubral-pallidoluysian atrophy, Machado-Joseph disease, and spinocerebellar ataxia type 1. Ann Neurol 1997; 41(4): 505-11.

Hensman Moss DJH, Pardinas AF, Langbehn D, Lo K, Leavitt BR, Roos R*, et al.* Identification of genetic variants associated with Huntington's disease progression: a genome-wide association study. The Lancet Neurology 2017.

Huang SR, Wu YT, Jao CW, Soong BW, Lirng JF, Wu HM*, et al.* CAG repeat length does not associate with the rate of cerebellar degeneration in spinocerebellar ataxia type 3. Neuroimage Clin 2017; 13: 97-105.

Jacobi H, du Montcel ST, Bauer P, Giunti P, Cook A, Labrum R*, et al.* Long-term disease progression in spinocerebellar ataxia types 1, 2, 3, and 6: a longitudinal cohort study. The Lancet Neurology 2015; 14(11): 1101-8.

Kennedy L, Evans E, Chen CM, Craven L, Detloff PJ, Ennis M*, et al.* Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis. Hum Mol Genet 2003; 12(24): 3359–67.

Lee J-M, Correia K, Loupe J, Kim K-H, Barker D, Hong EP*, et al.* Huntington's disease onset is determined by length of uninterrupted CAG, not encoded polyglutamine, and is modified by DNA maintenance mechanisms. 2019: 529768.

Limprasert P, Nouri N, Heyman RA, Nopparatana C, Kamonsilp M, Deininger PL*, et al.* Analysis of CAG repeat of the Machado-Joseph gene in human, chimpanzee and monkey populations: a variant nucleotide is associated with the number of CAG repeats. Human molecular genetics 1996; 5(2): 207-13.
Long A, Napierala JS, Polak U, Hauser L, Koeppen AH, Lynch DR*, et al.* Somatic instability of the expanded GAA repeats in Friedreich's ataxia. PLoS One 2017; 12(12): e0189990.
Maciel P, Gaspar C, DeStefano AL, Silveira I, Coutinho P, Radvany J*, et al.* Correlation between CAG repeat length and clinical features in Machado-Joseph disease. American journal of human genetics 1995; 57(1): 54-61.
McDaniel DO, Keats B, Vedanarayanan VV, Subramony SH. Sequence variation in GAA repeat expansions may cause differential phenotype display in Friedreich's ataxia. Movement disorders : official journal of the Movement Disorder Society 2001; 16(6): 1153-8.
Morales F, Couto JM, Higham CF, Hogg G, Cuenca P, Braida C*, et al.* Somatic instability of the expanded CTG triplet repeat in myotonic dystrophy type 1 is a heritable quantitative trait and modifier of disease severity. Hum Mol Genet 2012; 21: 3558–67.
Nakajima E, Orimo H, Ikejima M, Shimada T. Nine-bp repeat polymorphism in exon 1 of the hMSH3 gene. Jpn J Hum Genet 1995; 40(4): 343-5.
Nestor CE, Monckton DG. Correlation of inter-locus polyglutamine toxicity with CAG*CTG triplet repeat expandability and flanking genomic DNA GC content. PLoS One 2011; 6(12): e28260.
Pollard LM, Sharma R, Gomez M, Shah S, Delatycki MB, Pianese L*, et al.* Replication-mediated instability of the GAA triplet repeat mutation in Friedreich ataxia. Nucleic Acids Res 2004; 32(19): 5962-71.
Sakamoto N, Larson JE, Iyer RR, Montermini L, Pandolfo M, Wells RD. GGA*TCC-interrupted triplets in long GAA*TTC repeats inhibit the formation of triplex and sticky DNA structures, alleviate transcription inhibition, and reduce genetic instabilities. The Journal of biological chemistry 2001; 276(29): 27178-87.
Thornton CA, Johnson KJ, Moxley RT. Myotonic dystrophy patients have larger CTG expansions in skeletal muscle than in leukocytes. Ann Neurol 1994; 35(1): 104–7.
Watanabe M, Abe K, Aoki M, Kameya T, Kaneko J, Shoji M*, et al.* Analysis of CAG trinucleotide expansion associated with Machado-Joseph disease. J Neurol Sci 1996; 136(1-2): 101-7.